LISA: Reasoning Segmentation via Large Language Model

Year: 2023 | Citations: 682 | Authors: Xin Lai, Zhuotao Tian, Yukang Chen

---

Abstract

Although perception systems have made remarkable ad-vancements in recent years, they still rely on explicit human instruction or pre-defined categories to identify the target objects before executing visual recognition tasks. Such systems cannot actively reason and comprehend implicit user intention. In this work, we propose a new segmentation task - reasoning segmentation. The task is designed to output a segmentation mask given a complex and implicit query text. Furthermore, we establish a benchmark comprising over one thousand image-instruction-mask data samples, incorporating intricate reasoning and world knowledge for evaluation purposes. Finally, we present LISA: large Language Instructed Segmentation Assistant, which inherits the language generation capabilities of multimodal Large Language Models (LLMs) while also possessing the ability to produce segmentation masks. We expand the original vocabulary with a token and propose the embedding-as-mask paradigm to unlock the segmentation capability. Remarkably, LISA can handle cases involving complex rea-soning and world knowledge. Also, it demonstrates robust zero-shot capability when trained exclusively on reasoning-free datasets. In addition, fine-tuning the model with merely 239 reasoning segmentation data samples results in further performance enhancement. Both quantitative and qualitative experiments show our method effectively unlocks new reasoning segmentation capabilities for multimodal LLMs. Code, models, and data are available at github.com/dvlab-research/LISA.